JUN 2 6 1990

UNCLASSIFIED DTIC

DTIC FILE COPY (2)

ELECTE NOV 16 1990

AD-A229 432

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
...ITY ...NG SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 90 1121
6a. NAME OF PERFORMING ORGANIZATION The Pennsylvania State University	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS *(City, State, and ZIP Code)* Department of Computer Science University Park, PA 16802	7b. ADDRESS *(City, State, and ZIP Code)* AFOSR/NE, Bldg 410, Bolling AFB, DC 20332-6448.
8a. NAME OF FUNDING SPONSORING ORGANIZATION Air Force Office of Scientific Research	9. PROCUREMENT INSTRUMENT ID NUMBER AFOSR 87-0400
8c. ADDRESS *(City, State, and ZIP Code)* AFOSR/NE, Bldg 410, Bolling AFB, DC 20332-6448.	10. SOURCE OF FUNDING NUMBERS 2305B3

11. TITLE *(Include Security Classification)*
A Complexity Theory of Neural Networks (Unclassified)

12. PERSONAL AUTHOR(S)
I. Parberry, P. Berman, and G. Schnitger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT	15. PAGE COUNT
Annual Technical	89/3/15 to 90/3/14	1990, June 20	13

16. SUPPLEMENTARY NOTATION
The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of the Air Force Office of Scientific Research or the U.S. Government.

17. COSATI CODES	18. SUBJECT TERMS
	Neural networks, complexity theory, fault tolerance, learning. (RH)

19. ABSTRACT
Significant results have been obtained on the computational complexity of analog neural networks, and distributed voting. The computing power and learning algorithms for limited precision analog neural networks have been investigated. Lower bounds for constant depth, polynomial size analog neural networks, and a limited version of discrete neural networks have been obtained. The work on distributed voting has important applications for distributed computation in the presence of faults, and the management of replicated databases.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT unlimited	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Craig	22b. TELEPHONE 202-767-4931	22c. OFFICE SYMBOL NE

UNCLASSIFIED

A COMPLEXITY THEORY OF NEURAL NETWORKS

Ian Parberry, Piotr Berman, Georg Schnitger
Dept. of Computer Science
333 Whitmore Laboratory
Penn State University
University Park, Pa 16802.

14 April 1990

Annual Technical Report for Period March 15, 1989 to March 14, 1990

Prepared for

Dr. A. Craig
AFOSR/NE
Building 410
Bolling AFB DC 20332-6448

A Complexity Theory of Neural Networks: Annual Technical Report for Period March 15, 1989 to March 14, 1990

Ian Parberry
Piotr Berman
Georg Schnitger

Dept. of Computer Science
333 Whitmore Laboratory
Penn State University
University Park, Pa 16802.

1. Research Objectives

Complexity theory is the study of resource-bounded computation. The aim of this project is to study the amount of resources, in particular, time and hardware, used in neural network computations. Research will focus on four major topics:

1. The relative computing power of various neural network models.
2. Algorithms for neural network computations; upper-bounds, lower-bounds and completeness properties.
3. Fault-tolerant computation.
4. Learning.

2. Accomplishments

2.1. Research by I. Parberry

Much experimental neural network research involves analog neurons, which input real values, and output real values. However, whilst the theory of analog neural networks developed to date uses real numbers, experimental work is typically performed on digital computers. Surprisingly, the simulations bear out the theory, even thought the former is inherently discrete, and the latter inherently analog. Thus it appears the

neural networks are robust to precision. This is a particularly important trait, since it is impossible to fabricate analog hardware which has arbitrarily high precision. In particular, biological systems perform well with wetware which has analog behaviour, but only limited precision.

The Principal Investigator, Ian Parberry, with his Ph.D. student, Zoran Obradovic, undertook to investigate analog neural networks with limited precision. In digital simulations, the activation levels of the neurons are limited to some fixed number of values, k, which depends on the particular computer in use. The computational and learning complexity of limited precision analog neural networks was investigated, with a particular emphasis on how the number of neurons and running time scale with k, as well as the size of the problem being solved.

The key to the research was the demonstration in [11, 12] that limited precision analog neural networks with k activation levels are equivalent to discrete neural networks with k levels of activation, and k–1 thresholds, as opposed to the traditional single one. The computational complexity of these k-ary neural networks was studied in [11, 12], and the learning complexity in [13].

The work in [11,12] extends the traditional binary discrete neural network complexity theory (see [7]) to the new multi-level discrete case. The reader is referred to the journal papers for details. One typical result is that unlike the binary and ternary case, the threshold values for the k-ary case where $k > 3$ cannot be fixed. For example, in the binary case, the threshold can be made 0. In the ternary case, the two thresholds can be made 0 and 1. In the general case, no fixed thresholds will suffice. If k is restricted to grow only polynomially with the size of the problem being solved, then

polvnomial size, constant depth k-ary neural networks compute only functions from TC^0, the classical complexity class for binary neural networks. This implies that the superiority of analog neural networks over discrete binary ones can only confer a polynomial in size and a constant multiple in depth. However, that polynomial may still be significant.

The work in [13] extends the learning algorithms for the binary discrete neuron to the k-ary case. An efficient version of the Perceptron Learning Algorithm and Littlestone's Winnow Algorithm are given, proved correct, and analyzed.

Other work by Ian Parberry and his ex-student Peiyuan Yan involves progress on lower-bounds. Whilst it is extremely difficult to obtain exponential size lower-bounds on the size required by constant depth neural networks to compute certain functions, we have made small progress by restricting the power of the neurons. We [14] consider depth 2 circuits of mod–p and mod–q gates augmented with the limited use of AND and OR gates with small fan-in. We are able to show an exponential size lower-bound for certain depth-3 circuits of these gates for computing Boolean conjunction.

Ian Parberry is currently working on a book which describes the complexity theory of neural networks, based on the experience and some of the results obtained on this project.

3. Research by Georg Schnitger

We compare the computing power of threshold circuits and circuits composed of sigmoid gates. In a recent paper, Sontag (E. Sontag, "On the Recognition Capabilities of Feedforward Nets", Technical Report, SYCON - Rutgers Center for Systems and

Control, Rutgers University, 1990) considers the problem of deciding whether exactly one of two n-bit strings has a majority of ones. He shows that sigmoid circuits can solve this problem in depth 2 (one hidden layer) and constantly many gates. We show that constantly many threshold gates don't suffice.

This raises the question whether sigmoid circuits have a dramatically larger computing power than threshold circuits. Our answer is negative for the most commonly used sigmoid $s(x)=1/(1+\exp(-x))$: A sigmoid circuit computing with s gates and d hidden layers can be simulated by a threshold circuit with size poly(s + log W), and O(d) hidden layers. (Here W is an upper bound on the weights used by the sigmoid circuit.) We also establish the reverse, implying that, within a polynomial (for size) and a constant factor (for the number of hidden layers) sigmoid circuits and threshold circuits are equivalent.

The two above mentioned results apply to the case of binary input. Next we consider the case of real-valued input, a case modeling analog input. We show that the computing power of sigmoids dramatically decreases. In particular, we consider the problem of approximating trigonometric functions like sine(x) and cosine(x) for inputs x from the interval $[0,2^n]$. If x is given by an approximate representation in binary, sigmoids can approximate with poly(n) gates and constantly many hidden layers. If the real number x is input and weights of size at most $2^{poly(n)}$ are used, then poly(n) gates do not suffice if we would like to compute within constantly many layers. Consequently, it is advisable to supply special purpose hardware to allow a speedy conversion of real-valued input into (approximate) binary representation.

4. Research by Piotr Berman

My work in the period March 1989 to March 1990 was almost exclusively devoted to problems of fault tolerance in distributed systems, which in most instances involved various applications of majority voting.

With my current student, Mirjana Obradovic (who was supported by this grant) I am working on optimizing threshold gates; that is we would like to minimize the sum of the weights (assuming integer weights). When the weights are allowed to be large integers, then merely testing the equivalence of two gates is a co-NP complete problem, hence optimization cannot be feasible. However, when the sum of the weights of even one of the gates involved in the equivalence test is polynomial, then an equivalence test can in polynomial time return the confirmation of the equivalence or a counterexample. We have developed a heuristic which uses this equivalence test as follows. It maintains a list of examples for the given threshold gate, a list of proven inequalities of the form: this input should have the value of the target function at least as high as that input, and a short list of assumed inequalities. In turns, the heuristic constructs a minimal gate satisfying proven and assumed inequalities. If the latter gate is equivalent to the given one, the heuristic terminates, otherwise it uses the counterexample obtained to extend the list of proven inequalities or to modify the list of the assumed ones.

While this work is still in preparation, the partial results happened to have very interesting applications in the area of management of replicated data bases. (One of our papers was accepted for the presentation at 9th Symposium on Reliable Distributed Systems, in the Fall of this year). Here the subject is a data base in which data items

are replicated and distributed between some number of sites, which may improve the reliability (a failure of several data sites does not render a piece of date unreachable) and access (local rather than remote reads). A static scheme allows to perform a data base transaction dependent on the set of processors which can at a particular instance of time communicate with the originator of the transaction. In a voting scheme the sets of processors allowed to execute are characterized by a distribution of votes and a quorum threshold. We have characterised several important classes of systems in which voting scheme provides the optimal static scheme. Moreover, we introduced efficient and practical algorithms to compute the optimal distribution of votes. A part of our technique is an efficient test for the equivalence of threshold gates.

With my former student, Juan A. Garay (now at IBM T.J. Watson Research Center) I continued investigations on the Distributed Consensus problem. In this problem a group of processors has the task of reaching a common decision. Each processor has its initial option (typically, a 0/1 value), the common decision must be consistent with the initial option of one of the processors. There are two complications which make this problem non-trivial: the communication is conducted via bilateral links (so no 'public' vote is possible) and some of the processors are faulty. No assumptions whatsoever are placed on the behavior of the faulty processors, e.g. they could be controlled by an omniscient adversary.

The goal of our research was to provide solutions with better quality parameters than the previous ones. The parameters which we study are the following: the resiliency, i.e. the tolerated number of faulty processors, the number of communication rounds and the volume of communications. So far, we do not know any solution

which would be superior simultaneously in all these aspects. We found a solution which uses 1 bit messages, and has asymptotically optimal resiliency (3/4 of the optimum) and number of rounds (2 times optimum). In collaboration with K.J. Perry of IBM Watson we found a solution which has the optimal resiliency, while the message size is limited to 2 bits and the number of exchange rounds is 3 times larger than the optimal one. In both cases we can reduce substantially the number of rounds by increasing message size to a higher constant (this is quite important in practice, since the cost of sending one-page message and one-bit message is usually the same). Both protocols have the form of a simple sequence of votes, in the second protocol there is a possibility of casting an undecided vote (hence 2 bits in a message, rather then 1). These result and their applications are the subject of the conference presentations at ICALP and FOCS, as well as of the paper submitted to Journal of ACM.

Another group of results concerned protocols with optimal (rather than near optimal) number of rounds and relatively small (so-called polynomial) message size. One of these result was presented at FOCS and is the subject of the paper invited to the journal of Mathematical Systems Theory. Another is the subject of the paper submitted to FOCS. While these result are also based on voting, the votes are nested recursively, which easily leads to a huge message size. The techniques developed by me and Garay allow to avoid participation in most of possible votes, hereby reducing the message size. One important aspects of these techniques are the rules which allow to identify quickly the faulty processors that 'harm' the computation. We define a set of computationally easy rules of inference which allow to identify faulty processors and to deduct avoidable votes.

The experience gained in the work on Distributed Agreement allowed me to obtain some interesting results on fault diagnosis for multiprocessor distributed system (in cooperation with Andrzej Pelc of the University of Quebec; these result are presented at IEEE FTCS 20 conference). In the fault diagnosis model we assume that the faulty processors communicate and compute unreliably, but they can be detected by their network neighbors with some probability; moreover faulty processors form a random subset of the system. The previous diagnosis technique was based on a simple threshold: the processors are diagnosed to be faulty based on the number of 'failed tests' (a good processor may fail a test, if the latter is 'administered' by a faulty one). We have shown that the quality of diagnosis improves substantially if we form a graph of processors, and solve a maximum independent set problem for this graph (an arc is introduced between two processors whenever one claims that the other has failed its test). While the maximum independent set problem is in general not feasible, we have shown that it suffices to form a collection of very small graphs, and tackle them separatedly. Moreover, we have shown a scheme which allows to distribute the test result reliably through the system even with a very small number of connections (if we have n processors, then the number of links and tests is of the order $n \log n$, we have proven that this order of growth is sufficient and necessary).

5. Cumulative Publications

1. P. Berman and B. Roos, "Learning one-counter languages in polynomial time", To Appear in *Journal of Computer and System Sciences*.

2. P. Berman and G. Schnitger, "On the Approximation Complexity of the Independent Set Problem", Proc. 6th Annual Symposium on Theoretical Aspects of

Computer Science, pp. 256-268, 1989. To Appear in *Information and Computation.*

3. M. Obradovic, P. Berman and Z. Obradovic, ''Comparison of learning models for Boolean functions'', Technical Report CS-88-21, Dept. of Computer Science, Penn State University, May 1988.

4. B. Kalyanasundaram and G. Schnitger, ''Rounds versus time for the two person pebble game'', Proc. 6th Annual Symposium on Theoretical Aspects of Computer Science, pp. 517-529, 1989. To Appear in *Information and Computation.*

5. I. Parberry and G. Schnitger, ''Relating Boltzmann machines to conventional models of computation'', Neural Networks, Vol. 2, No. 1, pp. 59-67, 1989.

6. I. Parberry, ''A primer on the complexity theory of neural networks'', Technical Report CS-88-39, Dept. of Computer Science, Penn State University, October 1988.

7. I. Parberry, ''A primer on the complexity theory of neural networks'', in *Formal Methods in Artificial Intelligence: A Sourcebook,* pp. 217-268, ed. R. Banerji, North-Holland, 1990.

8. P. Berman, I. Parberry and G. Schnitger, ''The Complexity of Reliability and Constraint Satisfaction in Neural Networks'', Abstract, *Neural Networks,* Vol. 1, Supplement 1, 1988.

9. P. Berman, I. Parberry and G. Schnitger, ''On the complexity of reliability in neural networks'', In Preparation.

10. J. Chu and G. Schnitger, ''The Communication Complexity of Several Problems in Matrix Computation'', Proc. 1st ACM Symposium on Parallel Algorithms and

Architectures, 1989.

11. I. Parberry and Z. Obradovic, ''Analog Neural Networks of Limited Precision I: Computing with Multilinear Threshold Functions'', Advances in Neural Information Processing Systems 2 (Proceedings of the 1989 IEEE Conference on Neural Information Processing Systems), pp. 702-709, Morgan Kaufmann, 1990.

12. I. Parberry and Z. Obradovic, ''Computing with Limited Precision Analog Neural Networks'', Submitted to *Journal of Computer and Systems Sciences.*

13. I. Parberry and Z. Obradovic, ''Learning with Discrete Multi-Valued Neurons'', To Appear in the Proceedings of the Seventh Annual Machine Learning Conference, 1990. Submitted to *Journal of Computer and Systems Sciences.*

14. P. Yan and I. Parberry, ''An Algebraic Lower Bound Technique for Small Depth Circuits of Mod-p and Mod-q Gates'', To be submitted to *Information and Computation.*

15. W. Maass and G.Schnitger, ''Threshold Gates versus Sigmoids'', In Preparation.

16. P. Berman and J. Garay, ''Asymptotically Optimal Distributed Consensus'', Proceedings of the Sixteenth ICALP, Springer-Verlag Lecture Notes in Computer Science 372:80-94 (July 1989).

17. P. Berman and J. Garay, ''Efficient Agreement on Bounded-Degree Networks'', Proceedings of the 1989 International Conference on Parallel Processing, pp. I188-I191, (August 1989).

18. P. Berman, J. Garay, and K. Perry, ''Towards Optimal Distributed Consensus'', Proceedings of the Thirtieth IEEE Symposium on Foundations of Computer Science, pp. 410-415 (October 1989).

19. P. Berman and G. Schnitger, "On the Performance of the Minimal Degree Heuristic for Gaussian Elimination", *SIAM Journal on Matrix Analysis and Its Applications,* 11:83-88 (January 1990).

20. P. Berman, J. Garay, and K. J. Perry. "Recursive Phase King Protocols for Distributed Consensus", Department of Computer Science Technical Report CS-89-24, The Pennsylvania State University (August 1989).

21. P. Berman and M. Obradovic, "Voting and Other Static Schemes for Managing Replicated Data", Department of Computer Science Technical Report CS-89-46, The Pennsylvania State University (December 1989).

22. P. Berman and J. A. Garay, "Improved Protocols for Distributed Consensus", Department of Computer Science Technical Report CS-90-15, The Pennsylvania State University (March 1990).

23. P. Berman and M. Obradovic, "A Practical approach for Computing Optimal View Assignments", Department of Computer Science Technical Report CS-90-18, The Pennsylvania State University (March 1990).

24. P. Berman and J. A. Garay, "Better Masking for Better Consensus", Department of Computer Science Technical Report CS-90-24, The Pennsylvania State University (April 1990).

25. P. Berman and P., A. Pelc, "Distributed Probabilistic Fault Diagnosis for Multiprocessor Systems", To Appear in the Proceedings of the Twentieth Annual International Symposium on Fault-Tolerant Computing, Newcastle, United Kingdom.

26. M. Obradovic and P. Berman, "Voting as the Optimal Static Pessimistic Scheme for Managing Replicated Data", Accepted for publication in the Proceedings of the Ninth Symposium on Reliable Distributed Systems (October 1990). Huntsville, Alabama.

27. P. Berman and J. Garay. "Asymptotically Optimal Distributed Consensus", Submitted to *Journal of the ACM.*

28. P. Berman and J. Garay, "Cloture Votes: $n/4$-Resilient, Polynomial Time Distributed Consensus in $t+1$ Rounds", To appear in the special issue of *Mathematical Systems Theory* (editor H.R. Strong) on the State of the Art of Distributed Agreement.

6. Personnel

Name	*Position*
Dr. Ian Parberry	Principal Investigator
Dr. Georg Schnitger	Co-Principal Investigator
Dr. Piotr Berman	Co-Principal Investigator
Mirjana Obradovic	Research Assistant
Zoran Obradovic	Research Assistant

7. Conference Presentations

I. Parberry and Z. Obradovic, "Analog Neural Networks of Limited Precision I: Computing with Multilinear Threshold Functions", IEEE Conference on Neural Information Processing Systems, Natural and Synthetic, Denver, CO, Nov. 1989.

J. Chu and G. Schnitger, "The Communication Complexity of Several Problems in Matrix Computation", 1st ACM Symposium on Parallel Algorithms and Architectures, Santa Fe, NM, June 1989.

P. Berman and J. Garay, ''Asymptotically Optimal Distributed Consensus'', Sixteenth International Colloquium on Automata, Languages, and Programming, Stresa, Italy, July 1989.

P. Berman and J. Garay, ''Efficient Agreement on Bounded-Degree Networks'', International Conference on Parallel Processing, St. Charles, Illinois, August 1989.

P. Berman, J. Garay, and K. Perry, ''Towards Optimal Distributed Consensus'', Thirtieth IEEE Symposium on Foundations of Computer Science, Research Triangle Park, NC, October 1989.